

Die Komplexität der Datafizierung: zur Herausforderung, digitale Spuren in ihrem Kontext zu analysieren

Breiter, Andreas; Hepp, Andreas

Erstveröffentlichung / Primary Publication

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Breiter, A., & Hepp, A. (2018). Die Komplexität der Datafizierung: zur Herausforderung, digitale Spuren in ihrem Kontext zu analysieren. In C. Katzenbach, C. Pentzold, S. Kannengießer, M. Adolf, & M. Taddicken (Hrsg.), *Neue Komplexitäten für Kommunikationsforschung und Medienanalyse: Analytische Zugänge und empirische Studien* (S. 27-48). Berlin <https://doi.org/10.17174/dcr.v4.2>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Empfohlene Zitierung: Breiter, A., & Hepp, A. (2018). Die Komplexität der Datafizierung: Zur Herausforderung, digitale Spuren in ihrem Kontext zu analysieren. In C. Katzenbach, C. Pentzold, S. Kannengießer, M. Adolf, & M. Taddicken (Hrsg.), *Neue Komplexitäten für Kommunikationsforschung und Medienanalyse: Analytische Zugänge und empirische Studien* (S. 27-48). doi: 10.17174/dcr.v4.2

Zusammenfassung: Ziel des Kapitels ist es zu diskutieren, wie sich online hinterlassene digitale Spuren in ihrer Komplexität auf eine Weise kontextualisieren lassen, dass man sie in ihrem sozialen Bedeutungsgehalt analysieren kann. Hierzu werden nach einer allgemeinen Diskussion der dabei bestehenden Herausforderungen am Beispiel von Schulinformationssystemen mögliche Strategien der sozialen Kontextualisierung von automatisch generierten Daten herausgearbeitet. Kernargument ist, dass für eine solche Kontextualisierung qualitative Daten notwendig sind und damit eine Triangulation von Analysen großer Datenmengen mit qualitativen Verfahren. Ausgehend von solchen Analysen werden im Fazit des Beitrags hieraus Linien zukünftiger methodischer Entwicklungen im Bereich der Forschung zu digitaler Datenspuren aufgezeigt.

Lizenz: Creative Commons Attribution 4.0 (CC-BY 4.0)

Andreas Breiter & Andreas Hepp

Die Komplexität der Datafizierung

Zur Herausforderung, digitale Spuren in
ihrem Kontext zu analysieren

1 Einleitung

Ein Kennzeichen der heutigen, tiefgreifenden Mediatisierung unserer sozialen Welt ist ihre fortschreitende Datafizierung. Gemeint ist damit, dass mit der Digitalisierung verschiedener Medien diese nicht mehr nur Mittel der Kommunikation sind, sondern zunehmend auch solche der Generierung von Daten. Was immer Nutzerinnen und Nutzer mit digitalen Medien tun, in einer hochgradig mediatisierten sozialen Welt hinterlassen sie „Fußabdrücke“ (Girardin et al., 2008) dieser digitalen Mediennutzung. Diese kumulieren zu „digitalen Spuren“ (Thatcher, 2014). Teilweise handeln Nutzerinnen und Nutzer dabei bewusst, beispielsweise beim Hochladen von Fotos auf bzw. beim Schreiben von Kommentaren in den „timelines“ digitaler Plattformen. Oftmals sind sie sich dessen aber auch nicht bewusst, und es geschieht als Nebenfolge medienbezogener Praxis. Dies ist beispielsweise beim Gebrauch von Suchmaschinen der Fall oder beim Lesen von Online-Zeitungen, wo nur einer beschränkten Zahl von Nutzerinnen und Nutzern die Reichweite ihrer digitalen Spuren und deren Verwendung etwa in der Werbeindustrie klar ist (Turow, 2011).

Digitale Spuren werden aber nicht nur von den Nutzerinnen und Nutzern selbst generiert, sondern ebenso von *anderen*, sobald sie mit Bezug auf diese online in-

teragieren. Beispiele dafür sind das Synchronisieren von Adressbüchern oder das Taggen von Bildern, Texten und weiteren digitalen Artefakten mit Namen anderer Nutzerinnen und Nutzer. Digitale Spuren beginnen heutzutage schon vor der Geburt. Ein Beispiel dafür ist die „Mediatisierung der Elternschaft“ (Damkjaer, 2015): Wenn die Schwangerschaft fortlaufend über Apps und Plattformen mit einem Kommunikationsfluss begleitet wird, findet nicht nur eine Konstruktion von „Elternschaft“ vor der Geburt statt, sondern es werden auch Datenspuren des werdenden Kindes generiert. In diesem Sinne gilt in Zeiten tiefgreifender Mediatisierung für Individuen, aber auch für Kollektive und Organisationen: „Wir können nicht keine digitalen Spuren hinterlassen“ (Merzeau, 2009, S. 4). Datafizierung trägt damit zu einer zunehmenden Komplexität der sozialen Welt bei, indem mit ihr eine neue Ebene der auf Software delegierten Sinnkonstruktion besteht: ‚Hinter‘ der Ebene der eigenen Praktiken und den schon nur schwer abschätzbaren Folgen des eigenen Handelns schleicht sich eine weitere, in ihren Prinzipien kaum für die Akteure erfassbare Ebene ein, nämlich die der automatisierten Konstruktion sozialer Wirklichkeit durch Algorithmen. Eine besondere Komplexität im Sinne der Unbestimmbarkeit ergibt sich dabei durch den Umstand, dass deren Konstruktionsprinzipien für die Akteure kaum identifizierbar und damit nicht abschätzbar sind.

In methodischer Hinsicht ist die Komplexität datafizierter Wirklichkeitskonstruktion ein mehrfaches Problem für die empirische Medien- und Kommunikationsforschung (siehe u. a. Katzenbach & Pentzold, 2017; Scholl & Loosen, 2018). Bestehende Forschung zu Datafizierung zeigt verschiedene Herausforderungen auf: In vielen Fällen werden die APIs, die Zugang zu entsprechenden Daten eröffnen, von Unternehmen in herausgehobenen Machtpositionen – Apple, Facebook, Google etc. – kontrolliert (Hogan, 2018). Aber selbst wenn dieser Zugang besteht, sieht man sich mit einem noch größeren Problem konfrontiert (Rogers, 2017): Wie lassen sich diese Daten in ihrer Komplexität auf eine Weise kontextualisieren, so dass sie sich in ihrem *sozialen* Bedeutungsgehalt analysieren lassen?

In diesem Kapitel wollen wir dieses zweite Problem vertiefend diskutieren. Zuerst einmal wollen wir uns detaillierter mit digitalen Spuren als einem Phänomen der Komplexität auseinandersetzen. Dann werden wir das Beispiel von Schulinformationssystemen aufgreifen, um mögliche Strategien der *sozialen* Kontextualisierung von automatisch generierten Daten zu diskutieren. Dieses Beispiel ist dabei bewusst sehr einfach ausgewählt, um unser Kernargument zu pointieren – nämlich, dass für eine solche Kontextualisierung qualitative Daten

notwendig sind und damit eine Triangulation von Analysen großer Datenmengen mit qualitativen Verfahren. Ausgehend hiervon möchten wir einige Schlussfolgerungen zu allgemeinen Herausforderungen einer solchen Art von Forschung ziehen. Wir hoffen, mit dieser exemplarischen, detaillierten Betrachtung mögliche Linien zukünftiger methodischer Entwicklungen aufzeigen zu können.

2 Digitale Spuren als Phänomen der Komplexität

Wenn man digitale Spuren in einer ersten Annäherung als Kumulation „digitaler Fußabdrücke“ versteht, die durch die Nutzung digitaler Endgeräte und Dienste hinterlassen werden, handelt es sich dabei um ein vergleichsweise neues Gebiet der Medien- und Kommunikationsforschung. Allerdings lässt sich eine Beschäftigung hiermit in einer längeren Diskussion darüber verorten, ob „neue“ Medien auch „neue“ Methoden ihrer Erforschung bedürfen (siehe bspw. Golding & Splichal, 2013; Hutchinson, 2016). Zusätzlich gilt es eine solche Auseinandersetzung in der weitergehenden Diskussion um „digital humanities“ und deren Methoden zu sehen (Baum & Stäcker, 2015; Gardiner & Musto, 2015). Als Phänomen haben digitale Spuren eine anspruchsvolle, aber auch kontroverse methodische Diskussion nach sich gezogen. Hierbei zeigt sich eine mehrfache Komplexität des Phänomens.

Digitale Spuren jenseits von Big Data

Zuerst einmal ist es wichtig im Blick zu haben, dass digitale Spuren nicht mit „Big Data“ gleichzusetzen sind. Indem Big Data ein „amorpher Catch-All-Ausdruck“ (Kitchin & McArdle, 2016) ist, provoziert er erhebliche Diskussionen um seine Tragfähigkeit in zwei Extremen. Während einige Forscherinnen und Forscher die Idee von Big Data umfassend kritisieren (Andrejevic, 2014; boyd & Crawford, 2012), erscheinen Big-Data-Analysen anderen als die Zukunft der empirischen Forschung schlechthin (Mayer-Schönberger & Cukier, 2013; Townsend, 2013). Uns geht es hier aber um eine andere Zugangsweise, auch wenn wir später nochmals auf einige implizite Annahmen dieser Diskussion zurückkommen werden. Der Ausgangspunkt unserer Argumentation ist der Folgende: Digitale Spuren sind eine Art von Daten, die sozial bedeutungsvoll werden, weil diese eine technologisch basierte Konstruktion sind, bei der ein bestimmter Akteur oder ein bestimmtes Akteurshandeln

oder ein durch Akteure entwickelter Algorithmus Bezugspunkt der Konstruktion sind. Dabei wird typischerweise an Individuen gedacht, es kann prinzipiell aber auch um „supra-individuelle Akteure“ (Schimank, 2010, S. 327-341) wie Kollektivitäten und Organisationen als auch um selbstlernende Systeme gehen. Erst durch solch technologisch basierte Konstruktionen des In-Beziehung-Setzens von Daten mit Akteuren der sozialen Welt werden diese zu bedeutungsvoller Information. Dies ist der Grund, warum Firmen (und andere Organisationen) ein solches Interesse an der Aggregation dieser Daten in Bezug auf „reale“ Menschen haben. Für den Zweck der empirischen Forschung ist es entsprechend ein zielführender Ausgangspunkt, *digitale Spuren als numerisch produzierte Korrelationen verschiedener Arten von Daten zu verstehen, die in einer digitalisierten Mediumgebung individuelle, kollektive oder korporative Akteure durch ihre Praktiken generieren*.¹ Die Komplexität von digitalen Spuren ist in der Vielfalt der Art der Produktion dieser Daten zu sehen, in der Vielfalt der bestehenden Korrelationen.

Diese Definition hat einen engen Bezug zur Kritik an einem naiven Verständnis von Big Data (Puschmann & Burgess, 2014). Das Versprechen von Big Data ist, dass mehr oder weniger große Mengen an datenbasierten Informationen durch automatisierte Prozeduren ausgewertet werden können und sich hierdurch zukünftige Entwicklungen vorhersagen lassen, ohne dass man zuvor definierte theoretische Vorannahmen machen muss. Wie es bekannte Repräsentanten von Big-Data-Analysen formulieren: „Wir brauchen nicht notwendigerweise eine zutreffende Hypothese für ein Phänomen, um beginnen zu können, unsere Welt zu verstehen.“ (Mayer-Schönberger & Cukier, 2013, S. 73). Oder, wie es in einem Untertitel eines sich sehr gut verkaufenden Praxisbuchs heißt (Marr, 2015), ginge es um den „intelligenten Gebrauch von Big Data, Analytik und Metrik, um bessere Entscheidungen zu treffen und Leistung zu verbessern“.²

- 1 Der Ausdruck „Spur“ (engl. „trace“) umfasst verschiedenen Bedeutungen und Implikationen (to trace, track, traceable, traceability, tracing, etc.) und kann sowohl ein isoliertes Objekt als auch eine Praxis oder einen Prozess bezeichnen, dessen „Spur“ erfasst wird (Reigeluth, 2014, S. 249; Serres, 2002, S. 1). Aufgrund dieser allgemeinen semantischen Fülle des Begriffs bestehen einige Ambiguitäten des Begriffs der digitalen Spuren, die wir mit der obigen Definition auflösen wollen.
- 2 Im Bereich von Bildung sind die auf großen Datenmengen basierenden „learning analytics“ (Ferguson, 2012; Papamitsiou & Economides, 2014) die neue Vision der auf Algorithmen basierten Kontrolle und des Managements individueller

Solch ein Ansatz reduziert die Komplexität des Phänomens der digitalen Spuren auf ein „Big Data Paradigma“, dem es um das „Management von Daten und deren Transformation in nutzbares und verkaufbares Wissen“ (Elmer et al., 2015, S. 3) geht. Aus Sicht empirischer Methoden der Sozialwissenschaften basieren solche Hoffnungen zum Teil auf dem, was wir als „Mythologie von Big Data“ bezeichnen können: „große Datensätze versprechen mehr Intelligenz und Wissen, das Einsichten eröffnen kann, die zuvor unmöglich waren, und das mit einer Aura der Wahrheit, Objektivität und Genauigkeit“ (boyd & Crawford, 2012, S. 2) umgeben ist. Dieser Art von „social analytics“ (Couldry et al., 2015) liegt eine „graduelle Normalisierung von Datafizierung“ (van Dijck, 2014, S. 198) als dem neuen Paradigma in Wissenschaft und Gesellschaft zugrunde. Genau an dieser Stelle gilt es vorsichtig zu sein: Forschende im Feld von Big Data „tendieren dazu, Behauptungen über die Natur der Daten sozialer Medien als natürliche Spuren und über Plattformen als neutrale Mittler zu wiederholen“ (van Dijck, 2014, S. 199). Die Idee ist, dass sobald die einfache Arbeit des Datensammelns abgeschlossen ist, die „Daten für sich selbst sprechen“ (Mosco, 2014, S. 180) würden. Komplexitätsreduktion durch große Datenmengen, das ist hier der Wunsch. Oder anders formuliert: Big Data wird als ein einfacher Weg gesehen, um die Komplexität unserer datafizierten sozialen Welt zu bewältigen. Dies erscheint uns als zu kurz gedacht. Wir möchten im Folgenden unsere Sichtweise ausführen.

Die Interpretation von digitalen Spuren

In der jüngeren wissenschaftlichen Diskussion wurden digitale Spuren und die damit verbundenen Möglichkeiten der Informationsgewinnung zu einem Ausgangspunkt der grundlegenden Kritik sozialwissenschaftlicher Methoden. Das Argument ist, dass mit fortschreitender Datafizierung Methoden der Sozialwissenschaften in eine „Krise“ kamen, weil digitale Spuren als eine weit angemessenere Quelle erschienen als die Arten von Daten, wie sie typischerweise in den Sozialwissenschaften erhoben werden (Savage & Burrows, 2007): Während

Lernprozesse geworden. Aber wie Perrotta and Williamson (2016) schreiben, wird bei solchen Analysen oftmals die Produktion der zugrundeliegenden Datenstrukturen und Algorithmen sowie deren Rückbezug auf die Konstruktion von Macht im sozialen Leben vernachlässigt.

quantitative Umfragen und qualitative Interviews ursprünglich als innovative Beiträge des methodisch adäquaten Beschreibens und Verstehens der sozialen Welt galten, entstand die Überlegung, dass in der heutigen Zeit aufgrund der fortschreitenden Datafizierung und der damit verfügbaren Datenquellen klassische Verfahren der Sozialwissenschaften einen wesentlich beschränkteren Zugang zu den Prozessen der Konstruktion der sozialen Welt liefern. Die gegenwärtig primär machtausübenden Organisationen – Firmen, Verwaltungen, Regierungs- und Bildungsinstitutionen – erhielten eine Vielzahl ihrer Informationen durch eine fortlaufende Beobachtung und Analyse der verschiedenen digitalen Spuren, die Menschen hinterlassen. Gegen solche Quellen erschiene jede akademische Forschung, die auf Umfragen und Interviews basiert, als unzulänglich. Viele etablierte Methoden würden so durch die jüngste Datafizierung unter Druck geraten, da sie keine angemessenen Antworten mehr auf die Probleme der sozialen Welt liefern würden, eine Problematik, die als „soziales Leben der Methoden“ (Savage, 2013, S. 5) gefasst wurde. Entsprechend sei es notwendig, die „sozialwissenschaftlichen Methoden neu zusammenzusetzen“ (Ruppert et al., 2013, S. 22). Eine breit diskutierte Schlussfolgerung hieraus ist, über neue Formen der Datenerhebung und Auswertung nachzudenken, die auf „digitalen Methoden“ (Rogers, 2013, S. 1, 13) basieren. Methoden wie Crawling, Scraping oder Data Mining verwenden digitale Spuren als Quellen der empirischen Forschung.

Einige Wissenschaftlerinnen und Wissenschaftler gehen sogar noch einen Schritt weiter und argumentieren, digitale Spuren würden es zum ersten Mal ermöglichen, einen *direkten* Zugang zu Prozessen sozialer Konstruktion zu erhalten. Das vielleicht bekannteste Beispiel für eine solche Position ist Bruno Latours Integration von digitalen Spuren in seinen Ansatz der Sozialforschung (siehe Latour, 2007). Eine „digitale Rückverfolgbarkeit“ („digital traceability“, Venturini & Latour, 2010, S. 6) wird dann als eine Möglichkeit der Analyse von sozialen Prozessen in situ angesehen: „Wenn man interessiert ist an der Konstruktion sozialer Phänomene, so impliziert dies die Rückverfolgung [tracking] eines jeden der involvierten Akteure und jede der Interaktionen zwischen diesen“ (Venturini & Latour, 2010, S. 5). Mit digitalen Spuren, so das Argument, bestehe nun ein direkter Zugang zu solchen Informationen, da diese Methoden es gestatten, Prozesse des „Assembling“ in dem Moment zu erfassen, in dem diese stattfinden (siehe Latour et al., 2012; Venturini, 2012).

Aus unserer Sicht missversteht eine solche Argumentation den Kern digitaler Spuren und ihrer Komplexität. Erstens hat diese Position das grundlegende Problem der Fehlinterpretation der sozialen Welt als „flach“ und entsprechend als rekonstruierbar allein durch eine Analyse der „Fußabdrücke“ digitaler Medien. Dies ist ein Ansatz, der die Komplexität der datafizierten sozialen Welt auf die Ontologie einer flachen Gesellschaft reduziert.³ Zweitens und noch grundlegender missversteht ein solcher Ansatz digitale Spuren als etwas Neutrales, das einen direkten Zugang zur Gesellschaft eröffnen würde. Jedoch sind digitale Spuren keine neutralen Phänomene; eher basieren sie auf technischen Prozeduren von *machthabenden Institutionen*: den Unternehmen, Verwaltungen, Einrichtungen usw., die diese Daten interessensgeleitet produzieren. *Machthabend* meint hier, dass diese Institutionen Organisationen sind, die die Chance haben, den Charakter und die Struktur der Daten und Metadaten zu definieren, wie auch verschiedene Zwecke ihrer Nutzung. Bei der Analyse digitaler Spuren greift man entsprechend – ob als Individuum (Selbstständige, Hacker etc.), Kollektivitäten oder Organisation – durch solche Institutionen *kontrolliert* auf *mit bestimmten Zweckorientierungen produzierte* und nicht neutrale Daten zu. Es gilt entsprechend, digitale Spuren als Indikatoren sozialer Wirklichkeit kritisch zu hinterfragen im Hinblick auf die jeweiligen Perspektiven und die zugrundeliegenden Interessen ihrer Produktion.

Daten sind also nicht einfach „Rohmaterialien“, die einen direkten Zugang zu einer Komplexität der datafizierten sozialen Welt eröffnen (Borgman, 2015). Im Gegenteil besteht das primäre methodische Problem für eine empirische Forschung zu digitalen Spuren darin, diese im *sozialen Sinne* bedeutungsvoll zu machen, das heißt Kausalzusammenhänge, Beziehungen und vor allem auch Machtverhältnisse *jenseits* einfacher Aggregation und Korrelation automatisierter Datensammlung aufzudecken. Folglich ist die methodische Herausforderung weniger eine automatisierte Analyse als Ausdruck einer „flachen“ Gesellschaft. Eher besteht die interpretative Herausforderung darin, digitale Spuren mit *weiteren Datenquellen* in Beziehung zu setzen, um solche Spuren zu validieren bzw. interpretierbar zu machen und so nutzbar zu machen für anspruchsvollere Er-

3 Hier wird das allgemeinere Problem der Idee einer sozialen Welt als Summe von Assemblagen repliziert (siehe für die Kritik an einem solchen Ansatz Couldry & Hepp, 2017, S. 57-78).

klärungen und Verfahren der Theoriebildung (siehe Crampton et al., 2013; Hepp, 2017; Lohmeier, 2014b). An dieser Stelle müssen wir sehr vorsichtig sein, mögliche Missverständnisse zu vermeiden. Wir teilen die Position, dass Kompetenzen in neuen Formen „digitaler Methoden“ (Rogers, 2013) und „automatisierter Analysen“ (Sommer et al., 2014) für eine Kommunikations- und Medienforschung notwendig sind, die auf dem Stand der Zeit sein und zu der Diskussion um Datafizierung beitragen möchte (Hepp, 2016, S. 234-237). Gleichzeitig stehen wir jedoch jedem Ansatz kritisch gegenüber, der digitale Daten als eine direkte Quelle für die Beschreibung der Komplexität von Gesellschaft begreift. Wir brauchen die Kombination solcher Daten mit weitergehenden Informationen über die Kontexte, die Gegenstand der jeweiligen Forschung sind.

3 Kontextbezogene Analyse von digitalen Spuren: das Beispiel einer schulischen Lernplattform

Die wesentliche Herausforderung bei der Analyse digitaler Spuren liegt darin, die Daten in Bezug auf die jeweilige Figuration – den Verflechtungszusammenhang der relevanten Akteure – so zu kontextualisieren, dass sie einerseits die Nutzerin bzw. den Nutzer berücksichtigen, die die Spuren produziert haben, und gleichzeitig andererseits die Daten als Ausdruck einer sozialen Konstruktion zu begreifen. Somit spielt es eine Rolle, wie die digitalen Spuren in einer kommunikativen Figuration auf die jeweilige Akteurskonstellation, den Relevanzrahmen und die Kommunikationsformen bezogen werden können, in denen und durch die sie produziert werden (Hepp & Hasebrink, 2017a; Hepp & Hasebrink, 2017b).

Als Illustration für ein solches Vorgehen haben wir mit einer Lernplattform ein computergestütztes Kollaborationswerkzeug (für eine Übersicht siehe Haake et al., 2000), wie es in Schulen zunehmend eingesetzt wird, gewählt.⁴ Lernplattformen definieren den Online-Raum, in dem Daten durch Mediennutzung als digitale Spuren produziert werden und dadurch eine soziale Realität konstruiert wird. Somit sind sie nicht nur Mittel zur Datensammlung, sondern Werkzeuge

4 Im Kontext von Lernanwendungen wird auch häufiger von Lernplattformen oder Lernmanagementsystemen gesprochen (Schulmeister, 2005).

für wirkungsmächtige Prozesse der Konstruktion von Lehrenden und Lernenden und damit der Schule insgesamt (Breiter & Jarke, 2016). Die Rahmensetzung für diese Konstruktionen erfolgt insbesondere durch die Betreiber der technischen Systeme, die mit deren Hilfe entsprechende Analysen realisieren. Die Art, wie Daten in den kommunikativen Praktiken in Schulen hergestellt werden, spielt dabei eine zentrale Rolle. So werden Noten gesammelt und ausgewertet, um Entscheidungen wie etwa über Versetzungen zu unterstützen. Mit Hilfe digitaler Spuren können so Lernpfade der Schülerinnen und Schüler rekonstruiert, aber auch die Kommunikation zwischen Lehrkräften und Eltern dokumentiert werden. Lernplattformen spielen hierbei eine zentrale Rolle und werden derzeit zu universellen Werkzeugen der Datensammlung, da sich alle Mitglieder der Schule über sie vernetzen. Daher werden nicht nur die Lern- und Lehrprozesse, sondern auch die Verwaltungsprozesse in Schulen darüber abgebildet, was zu einer weiteren Kumulation digitaler Spuren führt.

In einer von uns durchgeführten empirischen Studie in deutschen Gesamtschulen⁵ wurde das Forschungsziel verfolgt, auf Basis der Kommunikationsprozesse ihrer zentralen Akteure die Schule als soziale Organisation zu rekonstruieren. Fragestellung war, ob und wie sich bestimmte Formen und Inhalte der Kommunikation auf der Online-Plattform in der interpersonalen Kommunikation in der Schule wiederfinden und umgekehrt. Dahinter stand die Annahme, dass eine erhebliche Diskrepanz zwischen der Repräsentation der Schulorganisation auf der Online-Plattform und der wahrgenommenen Organisation durch die zentralen Akteure besteht. Hierfür wurden verschiedene Kommunikationsformen (online, face-to-face, papierbasiert) untersucht. Ein Teil der Untersuchung fokussierte das Verhältnis zwischen schulischer Binnenkommunikation unter Lehrkräften im Schulgebäude und der auf der Lernplattform. Dahinter stand die Hypothese, dass in beiden Fällen sehr ähnliche Kommunikationsprozesse, die gleichen Akteure und Inhalte zu finden seien und somit die Lernplattform ein „Abbild“ darstelle. Wer sich also face-to-face in einem engen Kommunikationsnetzwerk im Kollegium befindet, nutzt hierfür auch die Lernplattform. Um dies zu untersuchen, wurden sowohl die Kommunikationsprozesse im Lehrerzimmer, auf den Fluren usw. als auch die digitalen Spuren, die in der Lernplattform hinterlassen wurden, erfasst.

5 Gefördert durch die Deutsche Forschungsgemeinschaft (BR-2273/10-1). Unser Dank gilt Arne Hendrik Ruhe für die Datenanalyse und die Informationsvisualisierung.

Logfile-Analyse digitaler Spuren

Technisch gesehen protokolliert die in den Schulen eingesetzte Lernplattform – wie alle server-basierten Systeme – die Nutzerbewegungen im System in Logfiles. Diese enthalten typischerweise Daten über Aufruf, Zeiten, Verbleib usw. (Liu, 2011; Markov & Larose, 2007; Oliner et al., 2012; Suneetha & Krishnamoorthi, 2009), zumeist im Extended Common Logfile Format:

Abbildung 1: Beispiel für Einträge im Logfile

```
1.2.3.4 - [25/Aug/2016:12:15:33 +0100] „GET /index.php HTTP/1.1“ 200 23578  
- „Webbrowser (System etc.)“  
1.2.3.4 - [25/Aug/2016:12:15:47 +0100] „GET /page2.php HTTP/1.1“ 200 15789  
„http://www.domain.com/index.php“ „Webbrowser (System etc.)“
```

Mit der Hilfe derartiger Protokollierungen (wie in Abbildung 1 an einem fiktiven Beispiel dargestellt) ist es möglich, die Nutzerin bzw. den Nutzer durch die IP-Adresse (1.2.3.4) zu identifizieren.⁶ Neben der IP-Adresse wird die Kombination von Webbrowser und Betriebssystem erfasst. Durch diese Zuordnung lassen sich Nutzerinnen und Nutzer über ihre gesamte Bewegung im technischen Systemverfolgen, denn die Einträge erfassen immer auch die Adresse der zuvor und danach besuchten Seite (im Beispiel in Abbildung 1 startet die Nutzerin auf der Seite index.php, verbleibt 14 Sekunden und geht dann weiter zu page2.php durch Klicken eines Hyperlinks). Diese Klicks werden „Aktionen“ genannt. Durch geeignete Verfahren lassen sich so Verlaufsmuster für jede einzelne Nutzerin erfassen und rekonstruieren.

Grundsätzlich gibt es fünf Wege, um Logfile-Analysen als Auswertung digitaler Spuren durchzuführen: (a) eine Darstellung, welche Seiten häufiger als andere

6 Bei der Auswertung solcher Datenspuren ist man damit nicht nur mit forschungsethischen Fragen konfrontiert, sondern unmittelbar auch mit datenschutzrechtlichen. Wir werden auf diesen Punkt im Weiteren näher eingehen.

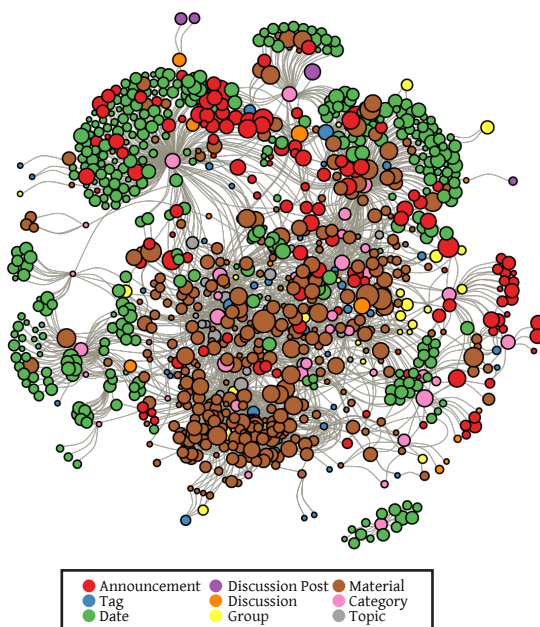
aufgerufen wurden und wie viele Nutzerinnen und Nutzer eine bestimmte Funktion verwendet haben; (b) ein Aufzeigen der Bewegungspfade; (c) eine Bündelung der Seitenbesuche in Cluster; (d) eine Netzwerkanalyse zur Identifikation von Verbindungen zwischen Nutzerinnen und Nutzer und (e) Mehrebenenanalysen unter Berücksichtigung verschiedener Organisations- oder Inhaltsebenen.

Logfile-Erhebungen sind nicht-reaktiv (Bortz & Döring, 2009): Alle Daten werden in der Applikation bzw. auf dem Server gesammelt und nicht durch die Nutzerinnen und Nutzer aktiv eingegeben. Somit handelt es sich um digitale Spuren, deren Erfassung für die User eher intransparent geschieht. Die Daten sind maschinenlesbar und können nicht nur entsprechend archiviert, sondern auch in Echtzeit erhoben werden. Neben diesen ausgewiesenen Vorteilen bestehen auch einige Nachteile bei der Datenerhebung. Zum einen geben die Datenspuren keine qualitativen Hinweise über das Nutzerverhalten. Ob eine bewusste Pause beim Durchlaufen eines Systems eingelegt wurde oder ob der nächste Schritt zu kompliziert war, kann aus den Logdaten nicht abgelesen werden. Wegen der fehlenden Reaktivität gibt es auch keine Angaben bspw. zum sozio-demografischen Hintergrund der Nutzerinnen und Nutzer. Und vor allem bestehen erhebliche Risiken hinsichtlich der Persönlichkeitsrechte, da Daten gesammelt werden, ohne dass der Nutzer bzw. die Nutzerin explizit ihre bzw. seine Einwilligung gegeben hat. Aus datenschutzrechtlicher Sicht ist dies nicht zulässig. Über die gespeicherten IP-Adressen lässt sich zudem in vielen Fällen rückverfolgen, wer die Nutzerinnen und Nutzer waren. Daher müssen aus forschungsethischen Gründen zusätzliche Vorkehrungen zur Anonymisierung getroffen und die Nutzerinnen und Nutzer so weit wie möglich über die Verfahrensweise informiert werden.

In unserem Anwendungsfall haben wir anonymisierte Daten aus einer Lernplattform extrahiert, die in einer großen Schule (über 100 Lehrkräfte und über 1.000 Schülerinnen und Schüler) genutzt wird. Das technische System steht vor allem den Lehrkräften und der Schulverwaltung zur Verfügung, um Kalender zu teilen, Foren zu unterhalten, Materialien bereitzustellen oder Räume zu buchen. Es ist auch außerhalb der Schule erreichbar und wird von einem privaten Anbieter gehostet.

Weitergehende interpretierende Schlussfolgerungen sind aus einer solchen Netzwerkvisualisierung allerdings nicht möglich. Daher wurden in einer Mehrebenenanalyse (Logfile-Analyse Variante e) andere Algorithmen eingesetzt, um weitere Zusammenhänge zunächst statistisch greifbar zu machen. Ziel war es, der

Abbildung 2: Daten aus der Lernplattform als Netzwerkgraph (Schulz & Breiter, 2013)



Frage näher zu kommen, ob sich die Kommunikationsformen und Akteurskonstellationen im Schulgebäude und auf der Lernplattform entsprechen. Die vorliegenden Logdaten umfassten einen Zeitraum von zwölf Monaten einschließlich der Schulferien. Insgesamt wurden 120.000 Zugriffe (Hits) von 138 Nutzerinnen und Nutzern aufgezeichnet. Im ersten Schritt wurden alle irrelevanten Einträge entfernt, wie die von Suchmaschinen oder Bots. Da manche Nutzerinnen und Nutzer ihre Sitzungen unterbrochen haben oder in einem neuen Fenster weitergeführt haben, mussten Algorithmen zur Vervollständigung von Pfaden eingesetzt werden. Durch beide Maßnahmen reduzierte sich die Zahl der verwertbaren Zugriffe auf 62.000. Für die 138 eindeutigen Nutzerinnen und Nutzer wurden insgesamt eine Zahl von 4.451 Besuchen (Visits) erfasst. Ein Besuch wird dabei als Folge von Hits eines eindeutigen Nutzers bzw. einer eindeutigen Nutzerin definiert, er endet nach einer Inaktivität von 30 Minuten (Catledge & Pitkow 1995).

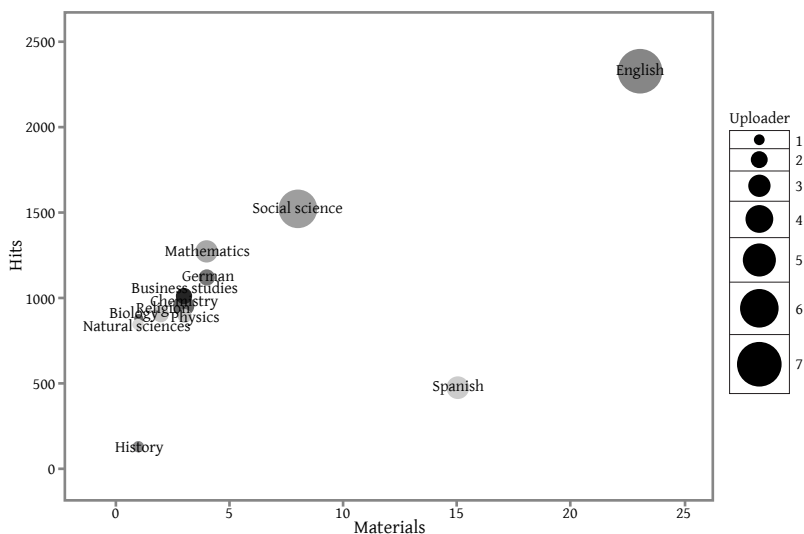
Abbildung 2 zeigt den zugehörigen Netzwerkgraphen (in obiger Terminologie eine Logfile-Analyse in der Variante d). Dabei werden Verbindungen als Kanten und Einträge (wie Termine) als Knoten repräsentiert, enge Verbindungen stärker gruppiert. Mit Hilfe dieser Visualisierungsform ist es möglich verschiedene Gruppen zu identifizieren, die weiter untersucht werden können. So finden sich im oberen Bereich des Graphen die Kategorien „Verschiedenes“, „Berichte“ und „Konferenzen“. Diese drei Kategorien sind vor allem durch das Item „Termine“ miteinander verlinkt, in selteneren Fällen durch die Items „Verabredungen“ oder „Materialien“. Die beiden letzteren Items werden hingegen häufiger von den Nutzerinnen und Nutzern abgerufen als Termine. Deutlich wird auch, dass die Kategorien stärker aufeinander und die Items auf die Kategorien verweisen als die Items gegenseitig aufeinander. Im unteren Bereich sind „Materialien“ repräsentiert, die sehr eng miteinander verbunden sind. Die darüber liegenden dickeren Knoten verweisen darauf, dass es eine starke Verlinkung zwischen diesen Items in der jeweiligen Kategorie gibt.

Der Scatterplot in Abbildung 3 (nächste Seite) zeigt nicht nur die Beziehung zwischen zwei Variablen (Materialien und Zugriffe), sondern über die Größe der Kreise in diesem Fall auch die Zahl der Personen an, die Materialien auf die Lernplattform hochgeladen haben. Es ist zu erkennen, dass der Spitzenreiter das Fach Englisch ist – sowohl bei den Zugriffen (2.300) als auch bei den Materialien (23). Spanisch zeichnet sich durch viele Materialien aus (15), aber nur 500 Zugriffe von drei aktiven Personen. Aber auch hier sind keine weitergehenden Aussagen dahingehend möglich, wie solche digitalen Spuren zu deuten sind bzw. wie sie sich auf Prozesse sozialer Konstruktion beziehen. An dieser Stelle braucht es nun qualitative kontextbezogene Daten, um die Ergebnisse zu interpretieren. Auf Basis der Logdaten wäre nur eine Spekulation möglich.

Kontextualisierung durch qualitative Verfahren

An diesem sehr knappen Beispiel einer Bewegungsanalyse auf einer Lernplattform mit Hilfe von Logdaten lässt sich somit bereits erkennen, dass die Interpretation von großen Datenmengen dieser Art nur durch kontextspezifisches Wissen möglich ist. Die zugrundeliegende Fragestellung nach vergleichbaren Kommunikationsmustern auf der Online-Plattform und im Schulgebäude lässt sich ohne solches Kontextwissen nicht beantworten, weil die Daten nicht ausreichend

Abbildung 3: Materialien und Zugriffe nach Fach (Quelle: Schulz & Breiter, 2013)



annotiert sind: Die digitalen Spuren alleine ermöglichen keine Aussagen zu den kommunikativen Praktiken und Akteurskonstellationen, durch die sie bzw. in denen sie generiert wurden. Die existierenden Algorithmen wie Clusteranalysen basieren auf Suche nach regelmäßigen und häufig wiederkehrenden Mustern. Um in unserem Beispiel Schulen als kommunikative Figurationen zu verstehen, braucht es weitere empirische Daten, die über die Logdaten hinausgehen. Wer sind die Akteure auf und außerhalb der Lernplattform? Was sind ihre Kommunikationspraktiken? Welche Rolle spielen diese Praktiken bei der Initiierung von Bewegungen auf den Lernplattformen und damit der Generierung digitaler Spuren? Und wie ist die Nutzung von Lernplattformen und damit die Generierung digitaler Spuren in ein weitergehendes Medienensemble eingebettet? Solche Fragen bleiben unserer Ansicht nach ohne qualitative Daten offen, da es hierbei auch um die Sinnorientierung der Akteure geht.

In unserem Fall wurde deswegen zusätzlich zur Analyse der Logdaten ein qualitatives Forschungsdesign basierend auf Interviews und teilnehmenden Beobachtungen eingesetzt (Welling et al., 2015). Über ein Schuljahr wurden Lehrkräfte im

Lehrerzimmer sowie in den Räumen der Jahrgangsstufenteams beobachtet. Mit Hilfe von Beobachtungsprotokollen wurden die Austauschbeziehungen, die Kommunikationsanlässe und die Medienrepertoires erhoben, kodiert und analysiert. Hinzu kamen Einzel- und Gruppeninterviews mit Lehrkräften, die mit Hilfe der dokumentarischen Methode (Bohnsack, 2000) ausgewertet wurden. In beiden Fällen war die Aneignung der Lernplattform durch die Lehrerinnen und Lehrer ein zentraler Bestandteil der Erhebung. Erst hierdurch konnte der Zusammenhang mit den identifizierten Mustern aus der Logfileanalyse hergestellt werden.

Unter Hinzunahme solcher Daten zeigte sich, dass die Aneignung der Lernplattform durch die Fachgruppe Spanisch vor allem darin begründet ist, dass es eine sehr aktive Lehrerin gab und die anderen Lehrkräfte an mehreren Schulstandorten unterrichten. Dadurch war die direkte Kommunikation wesentlich erschwert. Die Lernplattform wurde intensiv für das Zeitmanagement und den Austausch von Unterrichtsmaterialien genutzt. Bei der Fachgruppe Englisch zeigte die Analyse der Interviewaufzeichnungen und der Beobachtungsdaten eine langjährige Tradition im Austausch von Materialien. Aufgrund der Initiative einer Lehrkraft, die zuvor im Ausland unterrichtet hatte und dort selbstverständlich mit Lernplattformen arbeitete, war eine Unterstützung der Alltagskommunikation durch mediatisierte Formen umgesetzt worden. Es geht an dieser Stelle also darum, dass für eine vertiefende Interpretation der digitalen Spuren ein Wissen über die Akteurskonstellation der kommunikativen Figuration, in der diese entstehen, notwendig ist – in unserem Falle ein Wissen über die Akteurskonstellation der jeweiligen Fachgruppe.

4 Fazit: Herausforderungen für die Kontextualisierung digitaler Spuren

In seiner bewusst ausgewählten Einfachheit zeigt das Beispiel von Schulinformationssystemen, was es heißt, digitale Spuren zu kontextualisieren: Die Daten, die in den entsprechenden Systemen gesammelt werden, müssen mit weiteren, detaillierten Informationen in Verbindung gebracht werden, um diese Daten im sozialen Sinne deutbar zu machen. Erst dann werden diese Daten zu einer Quelle, um unsere zunehmend komplexe soziale Welt der Datafizierung zu beschreiben. Für eine empirische Forschung bestehen dabei drei Herausforderungen, die wir als grundlegend für jede sozialwissenschaftliche Analyse digitaler Spuren begreifen können.

Die erste Herausforderung ist, eine Möglichkeit zu finden, digitale Spuren mit Bezug auf eine bestimmte soziale Entität zu erfassen. Oftmals werden digitale Spuren ausschließlich als ein Phänomen einzelner Akteure begriffen, das heißt von Individuen hinterlassen, deren Spuren durch die Nutzung digitaler Medien und Dienste entstehen. Ein solcher Zugang ist adäquat für eine grundlegende Definition digitaler Spuren bzw. entsprechende Verfahren der Datengenerierung. Der Grund ist, dass es die einzelnen Nutzenden von Online-Systemen sind, die ihre „Fußspuren“ hinterlassen, die dann vom System gesammelt werden, häufig, weil der bzw. die Einzelne als Kunde bzw. Kundin interessiert. Unser Beispiel zeigt allerdings darüberhinausgehend, dass es wichtig ist, diese Individuen als soziale Akteure zu verstehen, deren Praktiken in die Figurationen weiterer institutioneller Kontexte und sozialer Gruppen eingebettet sind. Im Falle unseres Beispiels sind das die Organisation Schule und die verschiedenen Lehrerguppen. Erst, wenn man dies reflektiert, werden die Daten bedeutungsvoll. Figurationsanalytisch ausgedrückt müssen wir die gesamten kommunikativen Figurationen im Blick haben, in denen Individuen als „Urheber“ von digitalen Spuren agieren. Die Herausforderung an dieser Stelle ist es, eine Möglichkeit zu finden, die automatisiert generierten Daten mit einer weiter differenzierenden Analyse einer solchen Figuration zu verbinden. Als verbindendes Element lässt sich bei den Relevanzrahmen der Akteure der jeweiligen Figuration ansetzen, also letztlich bei der übergreifenden Sinnorientierung ihrer Praktiken. Im Falle unseres Beispiels ist dies die fachliche Handlungsorientierung der jeweiligen Lehrerguppen.

Eine zweite Herausforderung ist die Triangulation von quantitativen digitalen Methoden (im Falle unseres Beispiels der Logfile-Analyse) mit Formen der qualitativen Analyse, die die benötigte Kontextinformation liefern. Wie wir gesehen haben, ist die Kombination von automatisiert erhobenen Daten einerseits und qualitativen Interviews bzw. Gruppendiskussionen andererseits ein vielversprechender Weg. Ein solches triangulatives Vorgehen verweist aber wieder zurück auf die erste Herausforderung: Nur, wenn die Akteurskonstellation der erforschten Figuration bekannt ist, wird es möglich, solche Interviews und Gruppendiskussionen zu führen bzw. die Daten aufeinander zu beziehen. Diese eröffnen dann reichhaltige Daten, wobei diese Reichhaltigkeit gleichzeitig ein forschungsethisches Problem werden kann.

Entsprechend ist die Forschungsethik eine dritte Herausforderung. Komplexität aufzudecken heißt hier, dies auf ethisch angemessene Weise zu realisieren. Jeder Ansatz, der digitale Spuren auf solche Weise kontextualisiert, stellt eine Beziehung

her zwischen den (zum Teil von den Betroffenen ohne größeres Wissen darüber) hinterlassenen Daten und den weiteren Informationen zu einer Person. Die auf diese Weise generierten Erkenntnisse können sehr weitreichend sein – an einzelnen Punkten wesentlich weitreichender als das, was der Person möglicherweise selbst bewusst ist. Eine forschungsethische Konsequenz ist die Notwendigkeit, die betreffende Personen detailliert über die mit der Datenerhebung und Auswertung verbundenen Möglichkeiten zu informieren (und ihnen beispielsweise anzubieten, die gesammelten Informationen auf Wunsch offen zu legen). Eine andere Konsequenz ist, dass wir als Forschende sehr vorsichtig sein müssen, wie wir unsere Ergebnisse publizieren, da die Triangulation von digitalen Spuren mit weiteren Informationen die Möglichkeit schaffen kann, einzelne Personen zu isolieren und aufzufinden. Anonymisierung wird eine wichtige und zunehmend aufwändige Aufgabe, die weit über ein Unkenntlich-Machen der Realnamen hinausgeht.

Im Hinblick auf diese drei Herausforderungen ist es offensichtlich, dass die soziale Kontextualisierung von digitalen Spuren nicht einfach nur ein neues Feld der Kommunikations- und Medienforschung ist. Als ein solches macht sie es zusätzlich notwendig, auf veränderte Weise über die Beziehung von qualitativen und quantitativen Daten wie auch über (digitale) Forschungsethik nachzudenken.

Prof. Dr. Andreas Breiter ist Professor am Zentrum für Medien-, Kommunikations- und Informationsforschung an der Universität Bremen

Prof. Dr. Andreas Hepp ist Professor am Zentrum für Medien-, Kommunikations- und Informationsforschung an der Universität Bremen

Quellenverzeichnis

- Andrejevic, M. (2014). The Big Data Divide. *International Journal of Communication*, 8, 1673-1689.
- Baum, C., & Stäcker, T. (2015). Die Digital Humanities im deutschsprachigen Raum. Methoden – Theorien – Projekte. *Zeitschrift für digitale Geisteswissenschaften*, Sonderband 1. doi: 10.17175/sb001_023
- Bohnsack, R. (2010). *Rekonstruktive Sozialforschung – Einführung in qualitative Methoden*. Opladen & Farmington Hills: UTB.

- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge: MIT Press.
- Bortz, J., & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Bowker, G. C. (2014). The theory/data thing. *International Journal of Communication*, 8, 1795-1799.
- boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5), 662-679. doi: 10.1080/1369118X.2012.678878
- Breiter, A. (2014). Schools as mediatized organizations from a cross-cultural perspective. In A. Hepp & F. Krotz (Hrsg.), *Mediatized worlds* (S. 288-303). London: Palgrave.
- Breiter, A., & Jarke, J. (2016). Datafying education: How digital assessment practices reconfigure the organization of learning. *Communicative Figurations Working Paper No. 11*. Bremen.
- Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide web. *Computer Networks and ISDN Systems*, 27(6), 1065-1073.
- Couldry, N., Fotopoulou, A., & Dickens, L. (2015). Real social analytics: A contribution towards a phenomenology of a digital world. *British Journal of Sociology*, 67(1), 118-137. doi: 10.1111/1468-4446.12183
- Couldry, N., & Hepp, A. (2016). *The mediated construction of reality*. Cambridge: Polity Press.
- Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2), 130-139. doi: 10.1080/15230406.2013.777137
- Damkjaer, M. S. (2015). Becoming a parent in a digitized age: Facebook as an agent of change? – Performative, dialogical, and preservative Facebook strategies in the transition to parenthood. Paper presented at Nordmedia 2015, Kopenhagen, Dänemark, 13. August 2015.
- Earl, M. J. (2001). Knowledge management strategies: Towards a taxonomy. *Journal of Management Information Systems*, 18(1), 215-233.
- Elmer, G., Langlois, G., & Redden, J. (2015). Introduction. In G. Elmer, G. Langlois, & J. Redden (Hrsg.), *Compromised data: From social media to big data* (S. 1-13). London: Bloomsbury Academic.

- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-317. doi: 10.1504/IJTEL.2012.051816
- Gardiner, E., & Musto, R. G. (2015). *The digital humanities*. Cambridge: Cambridge University Press.
- Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE*, 7(4), 36-43. doi: 10.1109/MPRV.2008.71
- Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Hrsg.), *“Raw Data” is an Oxymoron* (S. 1-14). Cambridge: MIT Press.
- Golding, P., & Splichal, S. (2013). New Media, New Research Challenges an Introduction. *Javnost – The Public*, 20(2), 5-10.
- Hepp, A. (2016). Kommunikations- und Medienwissenschaft in datengetriebenen Zeiten. *Publizistik*, 61(3), 225-246. doi: 10.1007/s11616-016-0263-y
- Hepp, A. (2017). Theory and empirically based theory development. In J. Matthes (Hrsg.), *International encyclopedia of communication research methods*. Malden: Wiley-Blackwell.
- Hepp, A., & Hasebrink, U. (2017a). Kommunikative Figuretionen. Ein konzeptioneller Rahmen zur Erforschung kommunikativer Konstruktionsprozesse in Zeiten tiefgreifender Mediatisierung. *Medien & Kommunikationswissenschaft*, 65(2), 330-347. doi: 10.5771/1615-634X-2017-2-330
- Hepp, A., & Hasebrink, U. (2017b). Researching transforming communications in times of deep mediatization: A figurational approach. In A. Hepp, A. Breiter, & U. Hasebrink (Hrsg.), *Communicative Figurations: Transforming Communications in Times of Deep Mediatization* (S. 51-80). Palgrave Macmillan, London.
- Hogan, B. (2018). Social media giveth, social media taketh away: Facebook, friendships and APIs. *International Journal of Communication*, 12, 592-611.
- Hutchinson, J. (2016). An introduction to digital media research methods: how to research and the implications of new media data. *Communication Research and Practice*, 2, 1-6. doi: 10.1080/22041451.2016.1155307
- Ifenthaler, D. (2012). Learning Management System. In N. M. Seel (Hrsg.), *Encyclopedia of the Sciences of Learning* (S. 1925-1927). Boston: Springer.
- Katzenbach, C., & Pentzold, C. (2017). Theoriearbeit in der Kommunikationswissenschaft zwischen Komplexitätssteigerung und Komplexitätsreduzierung. *Medien & Kommunikationswissenschaft*, 65(3), 483-499. doi: 10.5771/1615-634X-2017-3-483

- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1). doi: 10.1177/2053951716631130
- Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). The whole is always smaller than its parts': a digital test of Gabriel Tardes' monads. *British Journal of Sociology*, 63(4), 590-615. doi: 10.1111/j.1468-4446.2012.01428.x
- Latour, B. (2007). Beware, your imagination leaves digital traces. *The Times Higher Literary Supplement*, 6. April 2007.
- Liu, B. (2011). *Web data mining: exploring hyperlinks, contents, and usage data*. Berlin: Springer.
- Lohmeier, C. (2014). The researcher and the never-ending field: Reconsidering big data and digital ethnography. In M. Hand & S. Hillyard (Hrsg.), *Studies in Qualitative Methodology* (S. 75-89). Bingley: Emerald.
- Markov, Z., & Larose, D. T. (2007). *Data mining the web: uncovering patterns in web content, structure and usage*. Hoboken: Wiley.
- Marr, B. (2015). *Big Data: Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance*. Chichester: Wiley.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: Die Revolution, die unser Leben verändern wird*. München: Redline.
- Merzeau, L. (2009). Présence numérique: les médiations de l'identité. *Les Enjeux de l'information et de la communication*, 1, 79-91.
- de Montjoye, Y., Radaelli, L., Singh, V. K., & Pentland, A. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221), 536-539. doi: 10.1126/science.1256297
- Mosco, V. (2014). *To the cloud: Big data in a turbulent world*. Boulder: Paradigm.
- Oliner, A., Ganapathi, A., & Xu, W. (2012). Advances and challenges in log analysis. *Communications of the ACM*, 55(2), 55-61. doi: 10.1145/2076450.2076466
- Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Journal of Educational Technology & Society*, 17(4), 49-64.
- Perrotta, C., & Williamson, B. (2016). The social life of Learning Analytics: cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, 43(1), 3-16. doi: 10.1080/17439884.2016.1182927

- Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication*, 8, 1690-1709.
- Reigeluth, T. B. (2014). Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society*, 12(2), 243-254.
- Rogers, R. (2013). *Digital Methods*. Cambridge: MIT Press
- Rogers, R. (2018). Otherwise Engaged: Social Media from Vanity Metrics to Critical Analytics. *International Journal of Communication*, 12, 450-472.
- Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, 30(4), 22-46. doi: 10.1177/0263276413484941
- Savage, M. (2013). The ‚social life of methods‘: A critical introduction. *Theory, Culture & Society*, 30(4), 3-21. doi: 10.1177/0263276413486160
- Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), 885-899. doi: 10.1177/0038038507080443
- Schimank, U. (2010). *Handeln und Strukturen. Einführung in die akteurstheoretische Soziologie*. Weinheim, Basel: Juventa.
- Scholl, A., & Loosen, W. (2018). Die Komplexität der Komplexitätsforschung: Theoretische Herkunft und Anwendung in der Journalismusforschung. In B. Dernbach, A. Godulla, & A. Sehl (Hrsg.), *Komplexität im Journalismus* (im Erscheinen). Wiesbaden: Springer VS.
- Schulmeister, R. (2005). *Lernplattformen für das virtuelle Lernen: Evaluation und Didaktik*. München: Oldenbourg.
- Schulz, A. H., & Breiter, A. (2013). Monitoring user patterns in school information systems using logfile analysis. In D. Passey, A. Breiter, & A. J. Visscher (Hrsg.), *Next Generation of Information Technology in Educational Management. ITEM 2012*. Berlin, Heidelberg: Springer.
- Serres, A. (2002). Quelle(s) problématique(s) de la trace? Abgerufen von http://poldoc.enssib.fr/sites/poldoc.enssib.fr/files/poldoc/documents/sic_00001397.pdf
- Smet, C. D., Bourgonjon J., Wever, B. D., Schellens, T., & Valcke, M. (2011). Researching instructional use and the technology acceptance of learning management systems by secondary school teachers. *Computers & Education*, 58(2), 688-696. doi: 10.1016/j.compedu.2011.09.013
- Sommer, K., Wettstein, M., Wirth, W., & Matthes, J. (Hrsg.). (2014). *Automatisierung in der Inhaltsanalyse*. Köln: Herbert von Halem.

- Suneetha, K. R., & Krishnamoorthi, R. (2009). Identifying user behavior by analyzing web server access log file. *International Journal of Computer Science and Network Security*, 9, 327-332.
- Thatcher, J. (2014). Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication*, 8, 1765-1783.
- Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. New York: WW Norton & Company.
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208.
- Venturini, T. (2012). Building on faults: how to represent controversies with digital methods. *Public Understanding of Science*, 21(7), 796-812. doi: 10.1177/0963662510387558
- Venturini, T., & Latour, B. (2010). The social fabric: Digital traces and qualitative methods. *Proceedings of Future En Seine 2009*. Cap Digital.
- Welling, S., Breiter, A., & Schulz, A. H. (2015). *Mediatisierte Organisationswelten in Schulen: Wie der Medienwandel die Kommunikation in den Schulen verändert*. Wiesbaden: Springer.